

3

PATENTS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Tomohiro SAKAI

Serial No. (unknown)

Filed herewith



LOOP DIAGNOSIS SYSTEM AND
METHOD FOR DISK ARRAY APPARATUSES

CLAIM FOR FOREIGN PRIORITY UNDER 35 U.S.C. 119
AND SUBMISSION OF PRIORITY DOCUMENT

Assistant Commissioner for Patents

Washington, D.C. 20231

Sir:

Attached hereto is a certified copy of applicant's corresponding patent application filed in Japan on June 21, 2000, under No. 185588/2000.

Applicant herewith claims the benefit of the priority filing date of the above-identified application for the above-entitled U.S. application under the provisions of 35 U.S.C. 119.

Respectfully submitted,

YOUNG & THOMPSON

By

Benoit Castel

Benoit Castel
Attorney for Applicant
Registration No. 35, 041
Customer No. 00466
745 South 23rd Street
Arlington, VA 22202
Telephone: 703/521-2297

June 19, 2001

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

JC979 U.S. PTO
09/883328
06/19/01


別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application: 2000年 6月21日

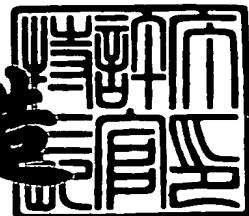
出願番号
Application Number: 特願2000-185588

出願人
Applicant(s): 日本電気株式会社

2001年 2月23日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特2001-3011535

【書類名】 特許願
【整理番号】 67000014
【提出日】 平成12年 6月21日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 3/06
【発明者】
【住所又は居所】 東京都港区芝五丁目7番1号 日本電気株式会社内
【氏名】 酒井 智弘
【特許出願人】
【識別番号】 000004237
【氏名又は名称】 日本電気株式会社
【代理人】
【識別番号】 100088812
【弁理士】
【氏名又は名称】 ▲柳▼川 信
【手数料の表示】
【予納台帳番号】 030982
【納付金額】 21,000円
【提出物件の目録】
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 ディスクアレイ装置のループ診断システム及びその方法

【特許請求の範囲】

【請求項1】 上位装置からの命令を実行するループを複数本有し複数個の記録媒体から構成される記録手段と、前記ループ各々の異常を検出するループ監視手段と、前記ループ監視手段における監視結果に応じて前記ループを制御するループ制御手段とを含むことを特徴とするディスクアレイ装置のループ診断システム。

【請求項2】 前記ループ制御手段は前記ループ監視手段にて特定ループの異常が検出された場合に前記特定ループにおける命令の実行を停止することを特徴とする請求項1記載のディスクアレイ装置のループ診断システム。

【請求項3】 前記ループ制御手段は前記特定ループにおける命令の実行を停止した後、前記特定ループ以外のループを使用して前記特定ループにおける命令の実行を代行することを特徴とする請求項2記載のディスクアレイ装置のループ診断システム。

【請求項4】 前記ループ制御手段は前記特定ループにおける命令の実行を代行した後、前記特定ループを診断することを特徴とする請求項3記載のディスクアレイ装置のループ診断システム。

【請求項5】 前記ループ制御手段は前記特定ループを診断した後、特定の前記記録媒体を前記特定ループから切離すことを特徴とする請求項4記載のディスクアレイ装置のループ診断システム。

【請求項6】 前記ループ制御手段は前記記録媒体を前記特定ループから切離した後、前記特定ループにおける命令の実行の停止を解除することを特徴とする請求項5記載のディスクアレイ装置のループ診断システム。

【請求項7】 前記ループ監視手段における監視結果と同等の情報を前記複数のループ以外から入力する保守用端末手段をさらに含み、前記ループ制御手段は前記保守用端末手段に入力された情報に応じて前記ループを制御することを特徴とする請求項1乃至6いずれかに記載のディスクアレイ装置のループ診断シス

テム。

【請求項8】 前記保守用端末手段は前記ループ制御手段から採取した情報を表示することを特徴とする請求項7記載のディスクアレイ装置のループ診断システム。

【請求項9】 上位装置からの命令を実行するループを複数本有し複数個の記録媒体から構成されるディスクアレイ装置のループ診断方法であって、

前記ループ各々の異常を検出するループ監視ステップと、前記ループ監視ステップにおける監視結果に応じて前記ループを制御するループ制御ステップとを含むことを特徴とするディスクアレイ装置のループ診断方法。

【請求項10】 前記ループ制御ステップは前記ループ監視ステップにて特定ループの異常が検出された場合に前記特定ループにおける命令の実行を停止することを特徴とする請求項9記載のディスクアレイ装置のループ診断方法。

【請求項11】 前記ループ制御ステップは前記特定ループにおける命令の実行を停止した後、前記特定ループ以外のループを使用して前記特定ループにおける命令の実行を代行することを特徴とする請求項10記載のディスクアレイ装置のループ診断方法。

【請求項12】 前記ループ制御ステップは前記特定ループにおける命令の実行を代行した後、前記特定ループを診断することを特徴とする請求項11記載のディスクアレイ装置のループ診断方法。

【請求項13】 前記ループ制御ステップは前記特定ループを診断した後、特定の前記記録媒体を前記特定ループから切離すことを特徴とする請求項12記載のディスクアレイ装置のループ診断方法。

【請求項14】 前記ループ制御ステップは前記記録媒体を前記特定ループから切離した後、前記特定ループにおける命令の実行の停止を解除することを特徴とする請求項13記載のディスクアレイ装置のループ診断方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明はディスクアレイ装置のループ診断システム及びその方法に関し、特に

FC-AL (Fibre Channel-Arbitrated Loop)
インターフェースのディスクを用いたディスクアレイ装置のループ診断システム及びその方法に関する。

【0002】

【従来の技術】

近年のIT (Information Technology) 環境の発展によりコンピュータ・システムにおける記憶装置の役割はますます大きくなってきており、記憶装置に対する高性能化、高信頼性化、大容量化などの要求も高まる一方である。ディスクアレイ装置においても、ホストとのインターフェースをFC-AL [Fibre Channel-Arbitrated Loop : ANSI (American National Standards Institute) X3. 272-199x Rev 5. 7 August 22, 1997にて規定されたインターフェース] 化したり、装置内に搭載するディスクのインターフェースをFC-AL化してそれらの要求に対応している。

【0003】

FC-AL (以下、ループと呼ぶ)においてリンクダウンなどのループ異常が発生してループ状態が乱れると、ループに接続されている他の正常なディスクに対する処理も影響を受け正常に処理することができなくなってしまう。ループを正常な状態に戻すためには故障部品をループから除去しなければならないが、間欠的なループ異常の場合には故障部品を特定するためには複雑なループ診断処理が必要であり、多数のディスクがループ接続された近年のディスクアレイ装置においては故障ディスクを除去するまでに要する時間は大い。

【0004】

この種のディスクアレイ装置の一例が、特開平11-353126号公報（以下、文献1と呼ぶ）に開示されている。文献1開示の技術は、ループを一旦切断してイニシエータホストとディスクアレイ装置とを1台ずつ順に接続して障害を調べる。そして、障害装置を特定するとその装置をループから切り離す。これにより、障害発生装置の特定が容易になり、かつ特定処理にかかる時間が短縮される、というものである。

【0005】

又、この種のディスクアレイ装置の他の例が、特開平11-305944号公報（以下、文献2と呼ぶ）及び特開平11-306644号公報（以下、文献3と呼ぶ）にも開示されている。文献2開示の技術は、リンクダウンの場合はランプを点灯させ、リンクアップの場合はランプを消灯させ、これによりリンクの状態を表示する、というものである。文献3開示の技術は、故障ディスクを切離した後、そのディスクを診断する、というものである。

【0006】

【発明が解決しようとする課題】

しかし、文献1開示の技術では、特定処理を実行している間、通常のディスク処理が中断するという欠点がある。又、ディスクアレイ装置の大容量化にともないループに接続されるディスク数が増大すればループ診断に要する時間はさらに増大し、通常のディスク処理が中断される時間もさらに増大するという欠点がある。一方、これらの欠点を解決する手段は前述の文献2及び3にも記載されていない。

【0007】

そこで本発明の目的は、特定処理を実行している間、通常のディスク処理が中断するのを回避することが可能なディスクアレイ装置のループ診断システム及びその方法を提供することにある。

【0008】

【課題を解決するための手段】

前記課題を解決するために本発明は、上位装置からの命令を実行するループを複数本有し複数個の記録媒体から構成される記録手段と、前記ループ各々の異常を検出するループ監視手段と、前記ループ監視手段における監視結果に応じて前記ループを制御するループ制御手段とを含むことを特徴とする。

【0009】

又、本発明による他の発明は、上位装置からの命令を実行するループを複数本有し複数個の記録媒体から構成されるディスクアレイ装置のループ診断方法であって、その方法は前記ループ各々の異常を検出するループ監視ステップと、前記

ループ監視ステップにおける監視結果に応じて前記ループを制御するループ制御ステップとを含むことを特徴とする。

【0010】

本発明及び本発明による他の発明によれば、記録手段は上位装置からの命令を実行するループを複数本有しており、特定ループで異常が発生してもその特定ループで実行すべき命令を他のループに代行させるため、特定処理を実行している間、通常のディスク処理が中断するのを回避することが可能となる。

【0011】

【発明の実施の形態】

以下、本発明の実施の形態について添付図面を参照しながら説明する。図1は本発明に係るディスクアレイ装置のループ診断システムの第1の実施の形態の構成図である。同図を参照すると、ディスクアレイ装置のループ診断システムの第1の実施の形態は、ディスクアレイ装置1で構成される。

【0012】

ディスクアレイ装置1は、ディスクユニット部5と、キャッシュ部6と、ディスク制御部A71と、ディスク制御部B72と、ホスト制御部A81と、ホスト制御部B82と、FC-AL状態監視部A131と、FC-AL状態監視部B132とから構成される。又、ディスク制御部A71、B72と、ディスク制御部A71、B72と、キャッシュ部6とは内部バス110により相互接続され、ディスク制御部A71、B72と、ディスク制御部A72、B72とは制御部間通信線100により相互接続される。

【0013】

さらに、ディスクユニット部5は、N(Nは正の整数)個のFC-ALディスク21~2Nと、ループ接続制御部A31と、ループ接続制御部B32と、FC-ALAループ41と、FC-ALBループ42とから構成される。

【0014】

ディスクアレイ装置1はホストインターフェースA121、B122を介してホストA91、B92と接続され、ホスト間とのコマンド受信、データ転送、ステータス応答などのホストサービスを行うホスト制御部A81、B82を有する。

ホストインターフェースの種別について制約はなく、FC-AL SCSI (small computer system interface) やパラレル SCSI など各種インターフェースを採用可能である。又、同図ではホストインターフェース数が 2 の場合を例示しているがインターフェース数についても限定するものではない。

【0015】

ホスト制御部 A81, B82 は内部バス 110 によりキャッシュ部 6、及びディスク制御部 A71, B72 に接続されており、お互いにデータの送受をおこなう。又、ホスト制御部 A81, B82 とディスク制御部 A71, B72 は、制御部間通信線 100 により、ディスク処理などの他の制御部に対する処理要求や、仮縛退状態通知や仮縛退解除通知などの制御部の状態通知などを行う。

【0016】

ディスクユニット部 5 は、FC-AL インタフェース・ポートを 2 ポート有する複数の FC-AL ディスク 21～2N と、ループ接続制御部 A31, B32 とから構成される。ループ接続制御部 A31, B32 は、FC-AL ディスク 21～2N の FC-AL ループ A41, B42 からの切り離し（バイパス）や、FC-AL ループ A41, B42 への接続を制御する。

【0017】

ディスク制御部 A71 はループ接続制御部 A31 を介して複数の FC-AL ディスク 21～2N の一方のポートのみで FC-AL ループ A41 を構成し、別のディスク制御部 B72 はループ接続制御部 B32 を介して FC-AL ディスク 21～2N のもう一方のポートのみで FC-AL ループ B42 を構成する。

【0018】

ディスク制御部 A71, B72 は、ホスト制御部 A81, B82 からの指示により、あるいはディスク制御部 A71, B72 自身の判断により、FC-AL ディスク 21～2N に対してリード (read) やライト (write) 等の各種ディスク処理を行う。又、ディスク制御部 A71, B72 は、ループ接続制御部 A31, B32 に指示することにより、任意の FC-AL ディスク 21～2N を FC-AL ループ A41, B42 からバイパスしたり、接続することが可能であ

る。FC-ALディスク21～2Nに対してどちらのディスク制御部A71, B72からもアクセスすることが可能である。

【0019】

ディスク制御部A71、B72と接続されているFC-AL状態監視部A131, B132は、ディスク制御部A(71、B72からディスク処理の実行結果を受け取り統計的に管理して、ループ異常の発生数あるいは発生率がしきい値を越えた場合にディスク制御部A71、B72に対して通知する。

【0020】

次に、本発明のディスクアレイ装置の動作について説明する。図1を参照すると、ホストA91, B92から命令を受け取ったホスト制御部A81, B82は、論理ディスク番号(LUN)、命令コード種別、及び論理ブロックアドレス(LBA)などの必要な情報を認識する。例えば、ホストA91, B92からのリード命令を受けたホスト制御部A81, B82は、キャッシュ部6に指定されたデータが存在すれば、直ちにキャッシュ部6からホストA91, B92にデータを転送する。

【0021】

キャッシュ部6にデータが存在しなければ、FC-ALディスク21～2Nから読み出したデータをキャッシュ部6に格納するようディスク制御部A71, B72に対して指示し、キャッシュ部6からホストA91, B92にデータを転送する。

【0022】

又、例えば、ホストA91, B92からのライト命令をうけたホスト制御部A81, B82は、ホストA91, B92から受け取ったデータをキャッシュ部6に格納する。このデータはホスト制御部A81, B82からディスク制御部A71, B72に対してディスクへの書き込み指示が出されることにより、あるいはディスク制御部A71, B72がディスクへの未書き込みデータがキャッシュ部6に存在することを検出することにより、ディスク制御部A71, B72によってFC-ALディスク21～2Nへ書き込まれる。

【0023】

ディスク処理の実行結果は、ディスク制御部A71, B72からFC-AL状態監視部A131, B132に通知され、FC-AL状態監視部A131, B132にて統計的に管理される。どのディスク制御部A71, B72がどのFC-ALディスク21～2Nに対して制御するかは、LUNやFC-ALディスク21～2N番号により割り振る、各ディスク制御部71, 72の稼働状態を調べ低負荷のディスク制御部A71, B72に動的に割り振る方式などがある。いずれにしても各ホスト制御部A81, B82と各ディスク制御部A71, B72間は制御部間通信線100によって連携を取り合うので、ホスト制御部A81, B82はどのディスク制御部A71, B72に対してディスク処理を指示するか、ディスク制御部A71, B72はどのFC-ALディスク21～2Nを制御すべきかの情報を共有している。

【0024】

次に、このディスクアレイ装置1の動作について図2及び図3を参照しながら説明する。図2はディスクのバイパス処理を説明するためのディスクアレイ装置1の部分構成図、図3はディスクアレイ装置1の動作を示すフローチャートである。

【0025】

なお、FC-ALループA41がループ異常となった場合を例として説明する。当然ながら、FC-ALループB42がループ異常となった場合も同様に動作する。例えば、いずれかのFC-ALディスク21～2Nの故障によりFC-ALループA41にてリンクダウン(link down)などのループ異常が間欠的に発生したとする(図3のS1参照)。FC-AL状態監視部A131がループ異常の発生数あるいは発生率がしきい値を越えたことを検出すると、ディスク制御部A71に対して通知する。

【0026】

この通知を受けたディスク制御部A71は、通常のディスク処理の実行を停止(この通常機能を停止している状態を仮縮退状態と呼ぶこととする)するとともに(図3のS2参照)、制御部間通信線100によりホスト制御部A81, B82及び別のディスク制御部B72に対して仮縮退状態に遷移したことを見出(仮

縮退通知) する (図3のS3参照)。

【0027】

この仮縮退通知を受けたディスク制御部B72は、最初にFC-ALディスク21～2Nのリセットを実行し、仮縮退ディスク制御部A71が放棄したディスク処理によりFC-ALディスク21～2N内に滞留している処理をキャンセルする。又、ディスク制御部B72は、すべてのFC-ALディスク21～2Nを対象に、ホスト制御部A81, B82から指示されたディスク処理、及びディスク制御部B72の判断によるディスク処理を行う (図3のS4参照)。

【0028】

又、この仮縮退通知を受けたホスト制御部A81, B82は、仮縮退したディスク制御部A71に対して要求していた未完了のディスク処理を代替のディスク制御部B72に対して再度要求する。又、ディスク制御部A71が仮縮退中は、新たなホストI/Oによるディスク処理についてすべて代替のディスク制御部B72に対して要求する。

【0029】

一方、仮縮退したディスク制御部A71は、FC-ALループA41に接続された複数のデバイスから故障部品を特定するための処理 (ループ診断) を行う (図3のS5参照)。

【0030】

ディスク制御部A71は、FC-AL A41, B42に接続されているすべてのFC-ALディスク21～2Nに対してループ診断のためのコマンド群を発行する。ディスク制御部A71は、ループ異常の発生状況を解析して異常の原因である被疑ディスクを特定し、例えば被疑部品がFC-ALディスク22ならばループ接続制御部A31に指示を出して、図2に示すようにFC-ALディスク22をFC-ALループA41からバイパス (bypass; 切離す) する (図3のS6参照)。

【0031】

そして、ループ診断を終えたディスク制御部A71は、制御部間通信線100によりホスト制御部A81, B82及び別のディスク制御部B72に対して仮縮

退状態を解除し通常状態に遷移したことを通知（仮縮退解除通知）する（図3のS7参照）。

【0032】

仮縮退解除したディスク制御部B71は、通常機能としてのディスク処理を再開する。仮縮退解除通知を受けたディスク制御部A72は、仮縮退解除したディスク制御部A71が受け持つディスクに対する処理を停止し、ループ診断により他方のループからバイパスされたディスクを含めた残りのディスクに対する処理を受け持つ。仮縮退解除通知を受けたホスト制御部A81, B82は、上記のディスクの受け持ちに応じたディスク制御部A71, B72に対してディスク処理を要求する。

【0033】

次に、第2の実施の形態について説明する。図4は本発明に係るディスクアレイ装置のループ診断システムの第2の実施の形態の構成図である。なお、第1の実施の形態と同様の構成部分については同一番号を付し、その説明を省略する（図1参照）。

【0034】

同図を参照すると、第2の実施の形態の構成が第1の実施の形態のそれと異なる点は、新たに保守用端末140を追加したことである。保守用端末140は、ディスク制御部A71, B72及びホスト制御部A81, B82と接続されている。

【0035】

保守用端末140からはディスク制御部A71, B72やホスト制御部A81, B82に対して命令を発行することができ、又、ディスク制御部A71, B72やホスト制御部A81, B82から採取した情報を保守用端末140に表示する。

【0036】

ディスク制御部A71, B72は、FC-AL状態監視部A131, B132からの指示だけでなく、保守用端末140からの指示によっても仮縮退状態に遷移してループ診断を行う。ループ異常の発生頻度が低くFC-AL状態監視部A

131, B132のしきい値を越えない場合や、定期保守の場合など、保守員の判断により保守用端末140からディスク制御部A71, B72に対してループの診断を起動することが可能である。

【0037】

又、第1の実施の形態で説明したようにループ異常によりFC-ALディスク22のFC-ALループA41側がバイパスされた状態となったとき、FC-ALループB42側のポート故障が発生したとしてもディスクアレイ構成であるためディスク縮退として継続運用することは可能であるが、早期に故障ディスクの交換を行い、FC-ALループA41とFL-ALループB42との2ループからの制御が可能な冗長構成に復帰させることが望ましい。保守用端末140は、バイパスされたFC-ALディスクが存在することを表示して、保守員に対してディスクの交換を促す。

【0038】

【発明の効果】

本発明によれば、上位装置からの命令を実行するループを複数本有し複数個の記録媒体から構成される記録手段と、前記ループ各々の異常を検出するループ監視手段と、前記ループ監視手段における監視結果に応じて前記ループを制御するループ制御手段とを含むため、特定処理を実行している間、通常のディスク処理が中断するのを回避することが可能となる。

【0039】

又、本発明による他の発明によれば、上位装置からの命令を実行するループを複数本有し複数個の記録媒体から構成されるディスクアレイ装置のループ診断方法であって、その方法は前記ループ各々の異常を検出するループ監視ステップと、前記ループ監視ステップにおける監視結果に応じて前記ループを制御するループ制御ステップとを含むため、上記本発明と同様の効果を奏する。

【0040】

より具体的には、本発明のディスクアレイ装置は、ディスクが接続されているFC-AL(ループ)において間欠的にループ異常が発生した場合、異常ループ側での通常のディスク処理を一時停止して故障部品を特定するためのループ診断

処理を行い、その間もホストI/Oなどによる通常のディスク処理は正常なループにより併行して行う。

【図4】

従って、ループ診断処理中もホストI/O処理が中断することなくループ診断時間を充分に確保でき精度の高い診断を行えるので、間欠的なループ異常であっても的確に故障部品を除去することができる。又、ループ診断時間に関する制約から解放されるので、ディスクの接続台数がさらに増大したり、診断精度を高めるためにループ診断処理のアルゴリズムを改良するためにループ診断時間が増加させることも可能である。

【図面の簡単な説明】

【図1】

本発明に係るディスクアレイ装置のループ診断システムの第1の実施の形態の構成図である。

【図2】

ディスクのバイパス処理を説明するためのディスクアレイ装置1の部分構成図である。

【図3】

ディスクアレイ装置1の動作を示すフローチャートである。

【図4】

本発明に係るディスクアレイ装置のループ診断システムの第2の実施の形態の構成図である。

【符号の説明】

- 1 ディスクアレイ装置
- 5 ディスクユニット部
- 6 キャッシュ部
- 2 1～2 N FC-ALディスク
- 3 1, 3 2 ループ接続制御部
- 4 1, 4 2 FC-ALループ
- 7 1, 7 2 ディスク制御部

特2000-185588

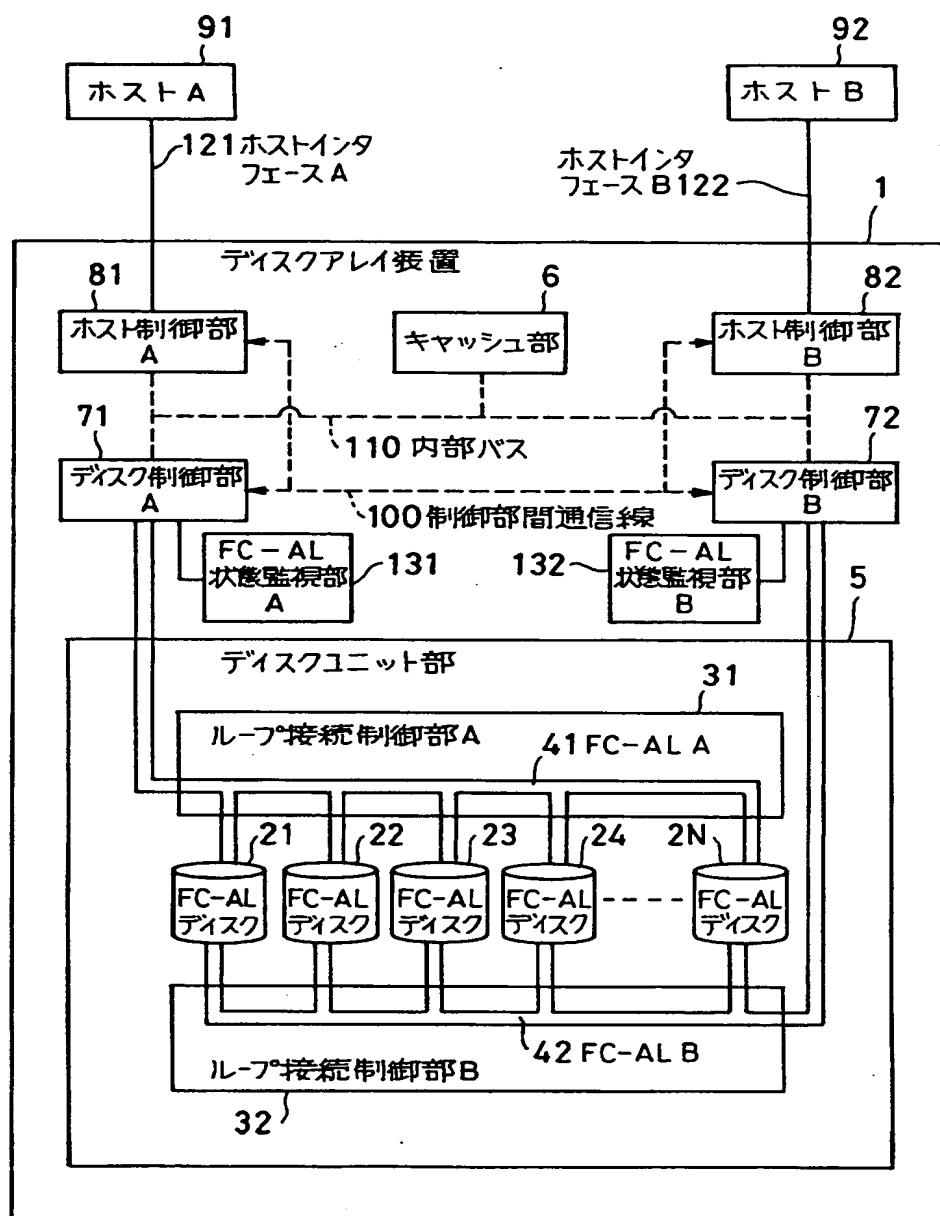
81, 82 ホスト制御部

131, 132 FC-AL状態監視部

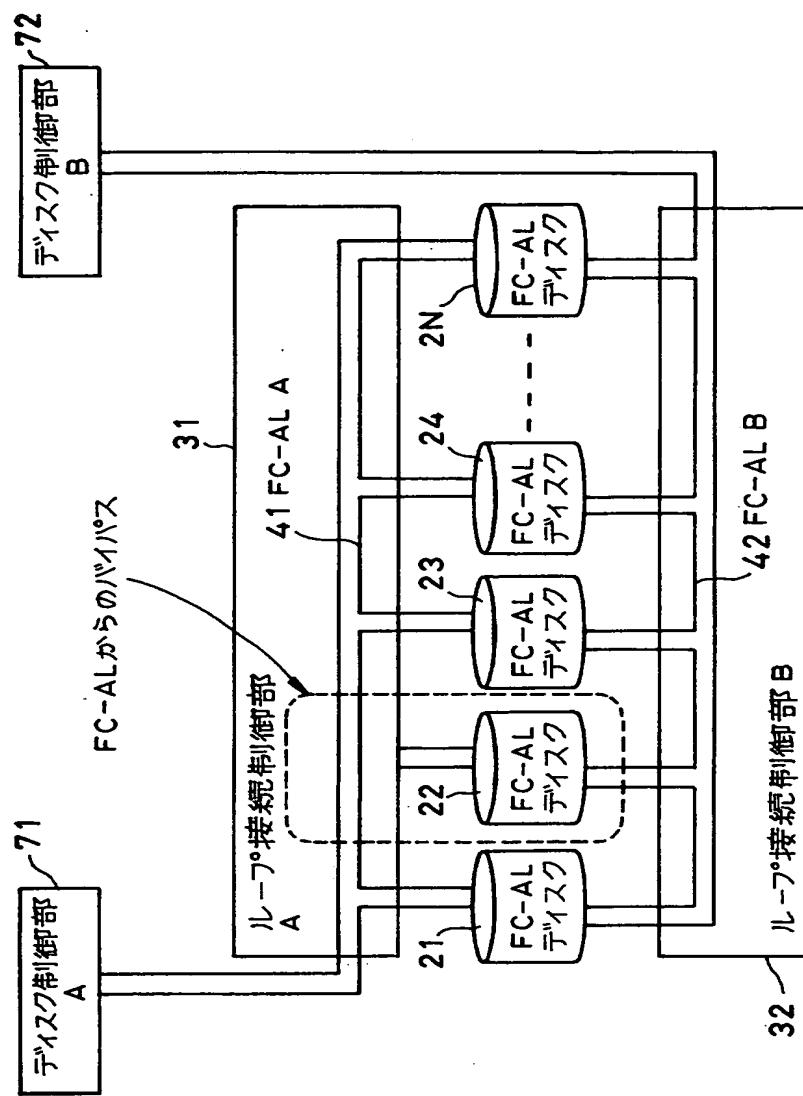
140 保守用端末

【書類名】 図面

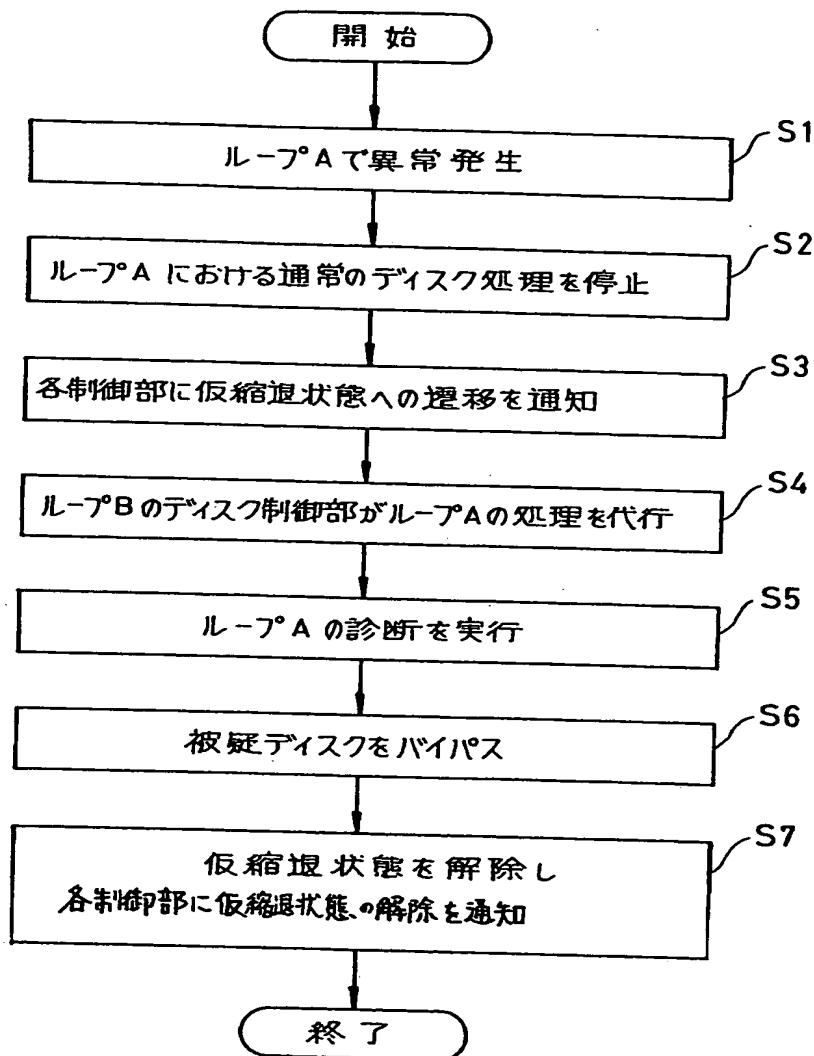
【図1】



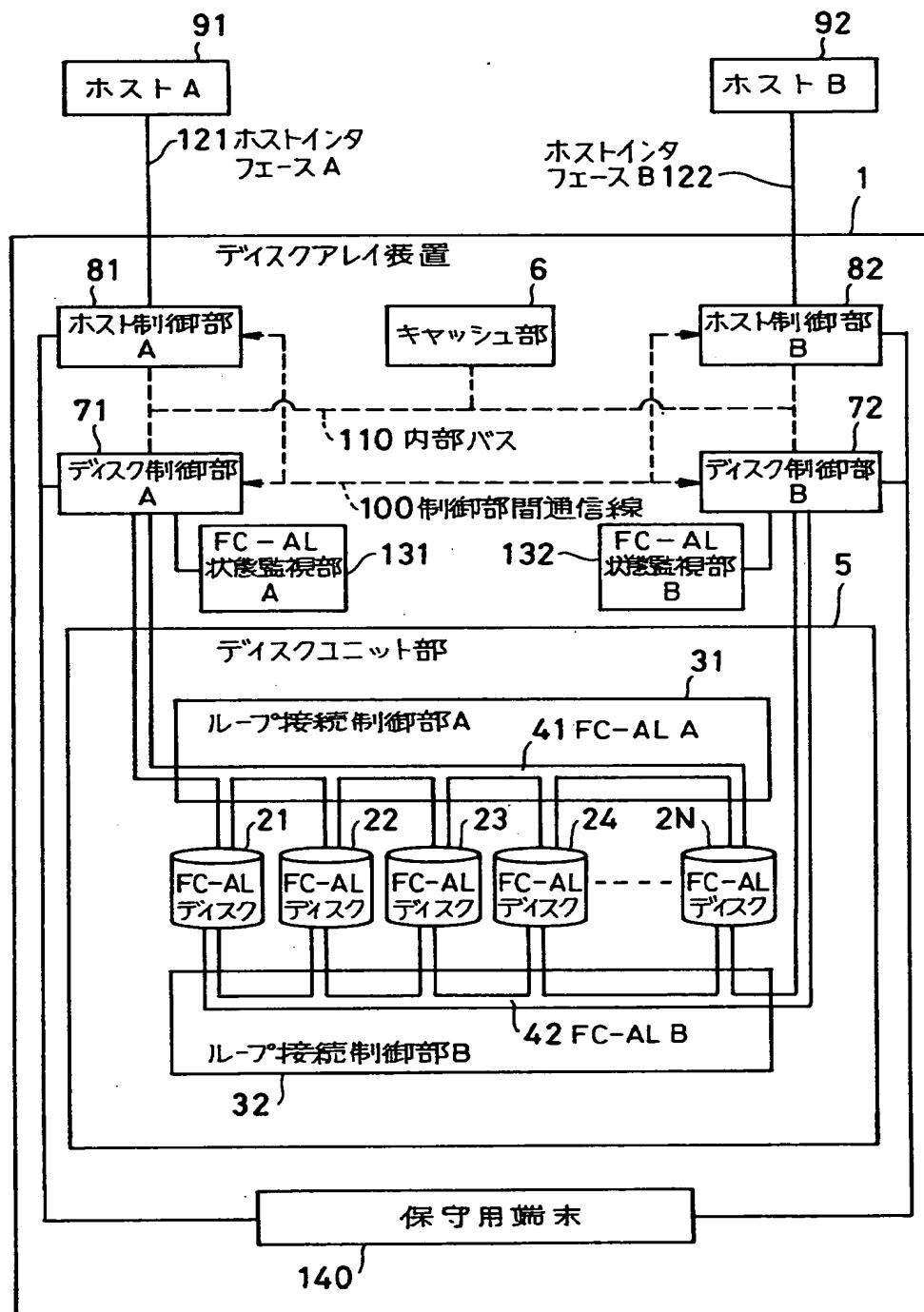
【図2】



【図3】



【図4】



【書類名】 要約書

【要約】

【課題】 特定処理を実行している間、通常のディスク処理が中断するのを回避する。

【解決手段】 ディスク21～2Nは2系統のループ41A, 42Bを有し、F
C-A L状態監視部A131でループA41の異常が検出されるとディスク制御
部A71はディスク制御部B72に未完了のディスク処理を要求する。ディスク
制御部B72はこの要求を受けループB42を用いてその未完了のディスク処理
を実行する。又、ホストからの新たな命令によるディスク処理もディスク制御部
B72が実行する。その後、ディスク制御部A71はループA41を診断する。

【選択図】 図1

出願人履歴情報

識別番号 [000004237]

1. 変更年月日 1990年 8月29日

[変更理由] 新規登録

住 所 東京都港区芝五丁目7番1号

氏 名 日本電気株式会社